DOCUMENT RESUME

ED 437 422 TM 030 576

AUTHOR Stake, Bob
TITLE Representing Quality in Evaluation.
PUB DATE 1999-04-00
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Change; Elementary Secondary Education; *Evaluation Methods; Evaluation Utilization; *Evaluators; *Program Evaluation; Urban Schools
IDENTIFIERS *Chicago Public Schools IL; Reform Efforts

ABSTRACT

Both quality and representation are constructs that are central to evaluation as a form of human discourse. Quality is perhaps the most essential idea in evaluation, and once quality is discerned, it must be conveyed, or represented to others. These constructs are considered in the context of the Chicago, Illinois public schools, an urban system that has undergone intense reform efforts. Quality, which does not mean the same thing to all people, does exist whenever someone recognizes it. The representation of quality depends on who has defined the quality as well as the planned use of the representation. Evaluators are caught up in the advocacies of schools and the provisions of the marketplace and in the compulsion of the accountability movement. One of the main responsibilities of evaluators is to represent the quality of their own industry accurately. They should not raise unattainable public expectation. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 437 422

REPRESENTING QUALITY IN EVALUATION

Bob Stake

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Bob Stake
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

TM030576

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in Montreal Canada
April 1999

BEST COPY AVAILABLE

Representing Quality in Evaluation[1]

Bob Stake

Ten years ago, U.S. Secretary of Education Bill Bennet declared Chicago schools the worst in the world. One year ago, President Bill Clinton declared the Chicago schools a standard for all to behold. What can we learn about quality from the representations of these two connoiseurs?

Quality and representation. We have two constructs here, both central to evaluation as a form of human discourse. Neither construct is easily defined.

First: Quality, perhaps the most essential idea in evaluation. To evaluate is to recognize quality. Evaluating is not *first*: the measuring of certain criteria. *And not first*:: facilitating decision-making. Evaluation is *first:* the search for goodness and badness, for merit and shortcoming, for quality.

The second elusive construct is Representation. In evaluation we not only need to *discern* quality but to *convey* the sense of quality to others. And then also, we need to comprehensively identify the evaluand, the thing about which quality is experienced. Thus, to our audiences, we represent twice; we represent quality and we represent the evaluand. It makes little sense to provide a sharp indication of quality if the object evaluated remains vague or misunderstood.

Original and derivative meanings. Let us dig into the origins of quality. Historically and personally, it comes out of comfort, out of happiness. Quality emerges not as a property of some object but out of experience with that object. No matter how complex the characterization of quality, its origins are rooted in experience.

As we express ourselves, we invent representations of quality. The resonance of quality becomes gradually less personal, more formalized. We invent criteria and standards, and they evolve. Back and forth, experience and verbalization interact. We speak of Grade A milk, not only because we like its taste but because we cherish the health we associate with eating well. The cherishing is basic; the verbal description is a derivative definition of quality. Grade A is a derivative status of milk.

A teacher becomes "Teacher of the Year." In summer school, a Chicago third-grader "fails the Iowas," again, and must repeat third grade, again. Michael Scriven tells us that, for wrist watches, Rolex is a standard of quality. The Rolex is a derived standard, useful for comparisons, only indirectly tied to personal experience. Sometimes we act as if these derivative definitions of quality are basic, that they get at some essence, that our experience is primitive and

[1] A presentation at the annual meeting of the American Educational Research Association, Montreal, Apr 11, 1999.

BEST COPY AVAILABLE

biased--and often it is, but definitions of quality need regularly to be traced back to how we are pleased, back to origins in experience.

Social Construction. As we rationalize our pleasures, as we come to understand our contexts, we see quality in an increasingly disciplined sense. Each of us develops a personal discipline. The discipline we develop for ourselves is not identical to the disciplines others develop. Quality becomes defined both ways, once through our own experiential participation, and again by beholding the standards of others.

For constructivists, quality doesn't exist until people declare it so. And people declare it when struck by the exquisite, when moved by the encounter. Probably simultaneously, we are aroused, we are emotionally involved, and we see it as something special, something aesthetic. The quality of teaching wasn't there until we saw it. Or until someone saw it and declared it.

At the same time we are developing our constructs of quality, others are developing theirs too. Much of what we construct, we construct together. Aesthetic quality is a social construction. The roots are in experience but the characterization, the representation of quality, often becomes far removed from experience.

Of course, some people do not experience what others experience, and thus a side of quality is hidden from them. The meaning of quality is formed by everybody, including those who do not experience it.

Disciplines of quality. As we come to understand our pleasures and the values around us, we see quality in an increasing aesthetic sense. When we encounter high quality in art and music, the exquisite; many are moved; many are thrilled; many are in awe. Many are not. The construction of merit is rooted in these experiences.

There is quality also in the mundane: in chairs, in pencils, even in paper clips. Are we thrilled by a paper clip? No, but we are dismayed if the cut-ending cuts our finger, if the clip doesn't close again after use, if the price of clips has doubled. Low quality is a mirror of our dismay.

The recognized disciplines of the arts and of all objects and performances of quality are not democratic institutions. It is not one person, one vote. As a result of some ordination, a certain ownership, various persuasions, a relatively few people exercise control over most formal disciplines.

Preserving independence. There are critiques, negotiations, political maneuvers, perhaps revelations. Agreements are struck and certain styles, sounds, movements and arrangements become standards. Ceaselessly, authority exists. The same is true for products

from cheeses to chiffoniers. Personal experience is sqeezed by the experience of others. The loudest and most eloquent advocate their criteria, and by their persuasions, maintain their roles as critic and judge.

Quality does not and should not reside alone in expert declamation. Standards need to be tempered and sometimes replaced by personal sensitivity. The reality of quality does not depend on the ability to explicate it.

Whether participant or spectator, whether producer or consumer, whether client or evaluator, the individual has to make a choice. The evaluator has to decide how much to rely on inside and outside voices. Sometimes the press of society makes it very difficult to exercise independence. Evaluators, like most people, cannot be satisfied with conventions of disciplined judgment that conflict with one's own personal discipline.

Correlates of quality. Many evaluators use a certain set of outcomes as their measure of quality. Some look pointedly for goal attainment, others for efficiency, still others for utility. In a context we know well, it becomes second nature for all of us to see quality upon encountering the searched-for characteristic. The Japanese have a saying, "It may be rotten but it is red snapper." We come to recognize the imprimatur of worth, the signs of success. We don't have to be joyous or aesthetically moved any more. Evaluators are expert at recognizing the surrogates for quality.

Evaluators sometimes indicate quality by referring to a formal standard. In our work it is usually a verbal standard. The standard in some way traces back to original encounters in which the standard setters were moved. All standards of quality thus are derivative. One can use a checklist of quality to evaluate a project, but the checklist needs to be validated against experience.

I have assured you I am not saying that one continues to be moved the same on subsequent encounters. We adapt to the presence of quality, knowing it increasingly relies on cognitive rather than conative grounds. Some new objects are so closely associated with objects already known that we attribute the quality of the one to the other. I do not say that a school is without quality if it fails to thrill us or repel us. I do say that school quality is often best thought of as a product of personal experience.

It will sometimes be useful to draw upon the experience of experts, of connoisseurs. It is important to know their perceptions of quality. Certainly, theirs are not the only important perceptions. Teachers, students, and other stakeholders have important perceptions. The usefulness of connoiseurs often is their ability to provide a language for comprehending the quality that others recognize but cannot communicate.

It will sometimes be useful to draw upon the skills of quantitative analysts. There is no obligation for the evaluator to aspire to some weighted synthesis of these various images.

But the analyst is an expert at identifying factors insufficiently discerned. With expert help, a panorama of awarenesses of quality is there for the evaluator to discern.

Egalitarian quality. Everybody constructs quality. I think we should take the egalitarian view that whosoever is moved, then there is quality. It doesn't mean that we have to admire everyone's quality, but we should respect it.

We grandparents observe the children performing. For those things we did not expect ours to be able to do, we are moved by the quality of it. The experts, the teachers see the performance less than the quality they are looking for. But it is quality, for us who are moved. I think evaluators need to find the audiences that are moved, and repelled, as well as to have the data scrutinized by blue ribbon reviewers.

It is quality for those who are moved, whether or not it is for the rest. Averages are not important. Standardization serves us poorly. We find life enriched by quality, however hidden it may be to everyone else. When we face standards that demean our enchantment, quality of life is lowered, not just for one, but for all.

Representation.

An evaluation report is a house of representatives. It is filled with representations of the evaluand and of quality. The evaluation report at the close of every World Bank overseas development project must explicitly say whether or not the work of the Bank was satisfactory or unsatisfactory. And by custom more than design, it at least briefly describes the irrigation, the training, the entrepreneur-support provided. It is important for the evaluator to let others know the nature of the evaluand perceived.

To represent is to stand for, to be a surrogate for, to present by typification, such as "words represent ideas." A name is a representation. I am Bob Stake. This is Linda Mabry. This is Elliot Eisner. Given our names alone, you know nothing of us. Some of you knew us well before. Jim Cullen introduced us, and perhaps added a bit more. But had Jim used the entire ninety minutes to find representations of us, you still would know but a snippet of who we are. It is the nature of a representation to be incomplete.

Our colleagues in quantitative research have relied overly much on the random sample for representing populations. Randomness assures precise indication of error variance and confidence intervals but does not assure representativeness. For many of the phenomena of education, there are no adequate representations. No case represents others, no legislator represents the people of her district, no member to the advisory committee represents people in some ways like them. Representativeness is a troubling concept in many aspects of evaluation work.

We sometimes seek symbolic representation that can be quickly comprehended, mere names, acronyms, logos, caricatures. And at other times, we seek representations that portray something of the complexity and the many contexts of the evaluand. We have colleagues such as Andy Porter, Michael Scriven and Patrick Grasso who, at least at times, urge the reduction of representation to the fewest possible indices. And we have colleagues, such as Linda, Elliot, Yvonna Lincoln and Jennifer Greene, who urge experiential accounts, narratives, stories to convey the sense of the evaluand as a living phenomenon. Still, they too, oversimplify. They try not to leave out the important qualities, but each reader has his or her own discipline.

Evaluation as interpretation. Measurement is easy. Interpreting measurement is difficult. Evaluators are challenged to give meaning to events, relationships, needs, and aspirations, far more complex than their words will convey. The representations created by the evaluator never completely mirror the things observed; all languages fail some, but some more than others. Analytic specification and operationalization are sometimes useful, not because they get closer to truth but because they can be grasped.

The reason for failure of experimental designs for evaluating educational reform is not so much the lack of a suitable control. Failure is assured because the "treatment" cannot be adequately represented. The "treatment" is never a single change element but a vast array of changes. There are so many "treatments" in the one experiment that good representation is impossible, leaving attribution confounded, and the evaluator bewildered, as to what might have caused any increase in quality of outcome.

The evaluator constantly seeks compromise between accuracy and comprehensibility. Both the case study and the performance indicator mislead. Both suggest meanings and precision not found in the object itself. Metaphors serve partly because they advertise their imperfect fit. Fallacies in representation need to be advertised. Lacking a statistical concept of standard error of measurement, we qualitative evaluators need to seek our own ways of informing the reader of the confidence that can be placed in our representations.

Deliberate misrepresentation. In these times, the most sophisticated representations are not those of the art or science but those of advertising. Lying is not a new invention, but the quality of lying has been greatly raised by Madison Avenue. Sound-bite campaigning. Image making.

What is the quality of our representations of educational quality. Are our executive summaries the equivalent of sound bites? Do we bury the trauma and immorality of the program in small print or impenitrable performance indicators? How frequently is the evaluator party to the lie?

I am going to show two very brief video clips The first is from a video prepared by the Chicago School District as part of its reporting of its Systemic Initiative Program to the National Science Foundation. Note the warm, friendly environment of the classrooms. Note the

interesting science gear the students have to work with. Note the emphasis on experiential education.

[show one minute video]

Using the skills of public relations specialists, the Chicago District is claiming that standards-based science is being taught in its schools. High quality science teaching **can** be found in a few classrooms, and almost every Chicago teacher creates a few moments of good science learning sometime during the year, but science is not taught as a regular subject in grade school and what is taught is more a matter of reading about space and dinosaurs than of learning science concepts. The video is a misrepresentation of the quality of science instruction and a misrepresentation as to the spread of science teaching.

A more representative look at a Chicago classroom occurred in a recent PBS Evening News item on Social Promotion. Inadvertently, PBS represented the classroom more accurately than the evaluation report.

[show 60 second video]

So here are the children, still nicely dressed, with caring teachers still, but learning in the usual recitation, stay in your seats, get ready for the Iowas curriculum.

The ethical standards of our profession insist on accurate representation. But Chicago badly needed the $5 million from the National Science Foundation, and will need more in the future. Can the evaluator blow the whistle? Can the evaluator avoid looking for those representations that will bring more support for the disadvantaged children of Chicago? It is a not a time of accuracy, but a time of advocacy.

But there is a kicker. I found out after I had written my first draft of this paper that the someone at the National Science Foundation asked for a video taped that showed the very best of student activity. The purpose was not to represent what Chicago is like but to support the claim that the Foundation has spent its funds well.

Representation of Quality. Evaluators will continue to be caught up not only in the advocacies of the schools and the provisions of the marketplace, but also in the compulsion of the accountability movement. In order to demonstrate the good work it is doing, every federal agency is caught up in the Government Performance Results Act. It is required to represent the quality of its impact by identifying aims, strategies, and outcomes. Here and elsewhere, the demand for representation far out-runs our ability to represent quality.

The qualitative "representation of misrepresentation," the counterpart to the standard error of measurement, is critical meta-evaluation, the periodic and climatic critical review of process and product, the challenge to description of evaluand and representation of

quality. These are procedures themselves lacking assurance of targetedness and relevance, but they are the best that evaluators have to offer.

One of the main responsibilities of evaluators is to accurately represent the quality of our own industry. We are not as good as our evaluation proposals and our promotional literature imply. We have helped raise public expectation of evaluation we cannot attain. Our searches for quality are too seldom of highest quality.